

Brat Disagreement Analysis Toolkit

Several reusable components have been developed for disagreement analysis in case of annotation tasks that use Brat. Please change your annotation schema in each file at the mentioned locations. Every entity and relation/event name has to be put into the configuration map.

[General Information]

Java 7 is needed for optimal utilization.

- The framework uses an external APIs that have to be specified in the build path of the project.
- Metamap WEB API (UMLS concept matching)
http://ii.nlm.nih.gov/Web_API/index.shtml
- Example collections for testing are provided in the folder (GoldenStandard_Iteration_12_S, GoldenStandard_Iteration_13_S). These can be visualized directly using Brat, as we provide the annotation schema too.

Inter-Annotator Agreement

Package: nih.org.bratqualityanalyzer.interannotatoragreement

- o DDIAgreementCalculator.java

Computes the inter-annotator-agreement between two sets of annotated files. Standard F-measure is used for individual entities.

Example output:

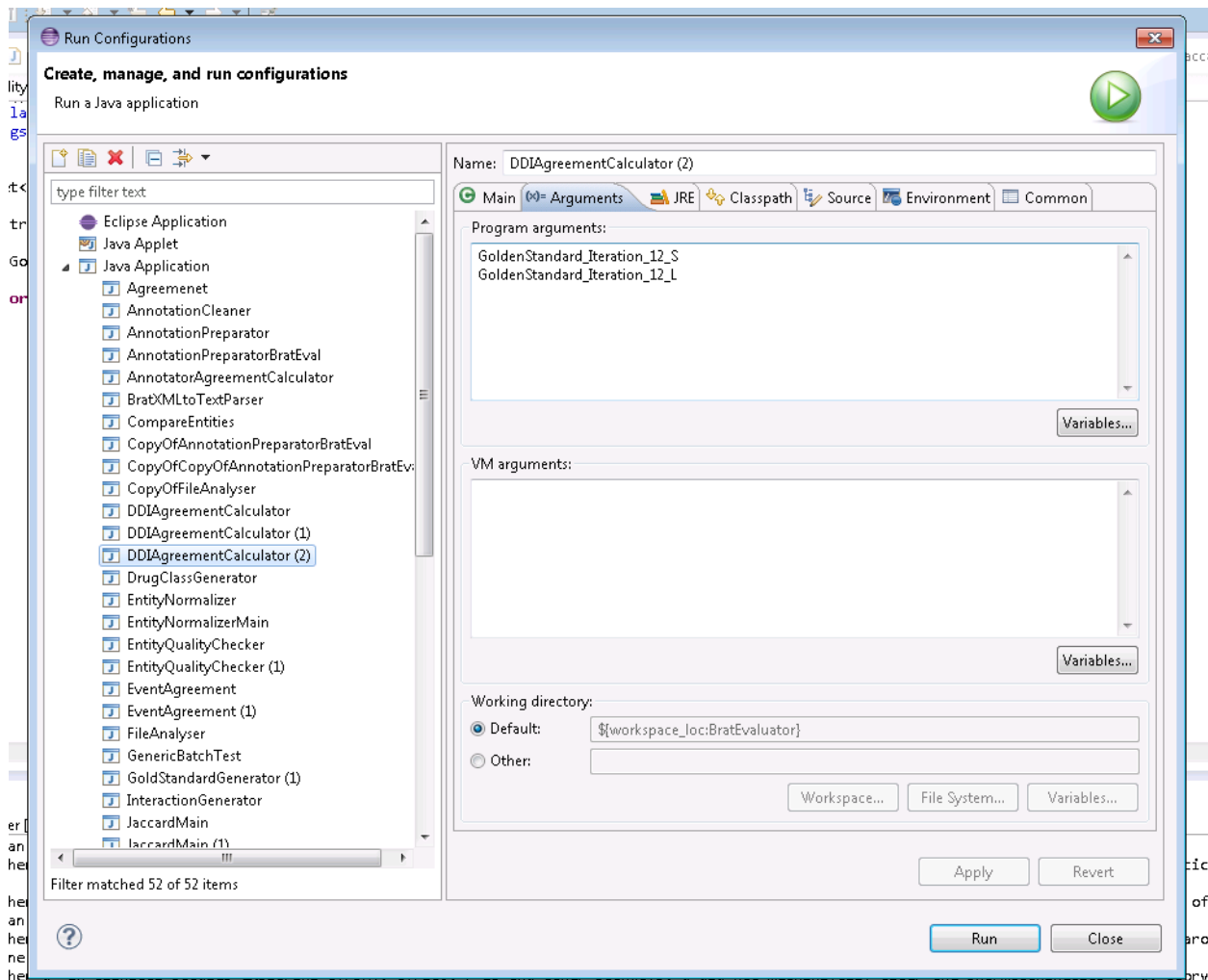
Differences between : .\GoldenStandard_Iteration_12_S\0742cf97-601b-4c13-a6c3-1a6a616ed292.ann File 2
 .\GoldenStandard_Iteration_12_L\0742cf97-601b-4c13-a6c3-1a6a616ed292.ann

Event Increase_Interaction	22	26	17
Drug_Class	20	19	19
Specific_Interaction	37	27	32
Caution_Interaction	14	9	8
Event Caution_Interaction	14	9	1
Substance	6	6	5
SPAN	20	11	7
Increase_Interaction	22	26	20
Decrease_Interaction	4	5	4
Event Specific_Interaction	37	27	12
hasDrugMember	8	8	8
hasClassMember	4	6	4
Drug	82	76	72
Event Decrease_Interaction	4	5	3

Expression	5	3	1
29 MISSING FROM FILE 2	Caution_Interaction	avoided	1200
29 MISSING FROM FILE 2	Drug	Verapamil	1180
29 MISSING FROM FILE 2	Drug_Class	beta-adrenergic blocker	1434
29 MISSING FROM FILE 2	Event	Caution_Interaction	avoided
hasContext:ventricular dysfunction hasPrecipitant:beta-adrenergic blocker hasObject:Verapamil			
29 MISSING FROM FILE 2	SPAN	ventricular dysfunction	1386
51 CHECK TEXT	hypotension	4848 Specific_Interaction	severe
hypotension 4841 Specific_Interaction			
51 MISSING FROM FILE 1	Coreference	Concomitant administration-	
->verapamil			
51 MISSING FROM FILE 1	Event	Specific_Interaction	severe
hypotension hasObject:Concomitant administration			
hasPrecipitant:quinidine			
51 MISSING FROM FILE 1	Expression	Concomitant administration	4763
51 MISSING FROM FILE 2	Coreference	whom-->8 patients	
51 MISSING FROM FILE 2	Event	Specific_Interaction	hypotension
hasPrecipitant:quinidine hasEvidence:3 of the 8 patients			
hasObject:verapamil			

Caution_Interaction|Common annotations:60|Annotator 1
only:15|Annotator 2 only:8|IAA:0.8392
Decrease_Interaction|Common annotations:17|Annotator 1
only:6|Annotator 2 only:6|IAA:0.7391
Drug|Common annotations:305|Annotator 1 only:53|Annotator 2
only:32|IAA:0.8777
Drug_Class|Common annotations:183|Annotator 1 only:28|Annotator 2
only:21|IAA:0.8819
Increase_Interaction|Common annotations:74|Annotator 1
only:5|Annotator 2 only:7|IAA:0.9250
SPAN|Common annotations:38|Annotator 1 only:28|Annotator 2
only:24|IAA:0.5938
Specific_Interaction|Common annotations:118|Annotator 1
only:26|Annotator 2 only:8|IAA:0.8741
Substance|Common annotations:12|Annotator 1 only:3|Annotator 2
only:1|IAA:0.8571
Global IAA:0.8562

- Usage: java [directory of annotated files of the first annotator] [directory of annotated files of the second annotator]
- Results are written into a file will be in the directory called [directory of annotated files of the second annotator]_out
- Screenshot for run configuration in Eclipse using test collections provided with the source code (GoldenStandard_Iteration_12*):

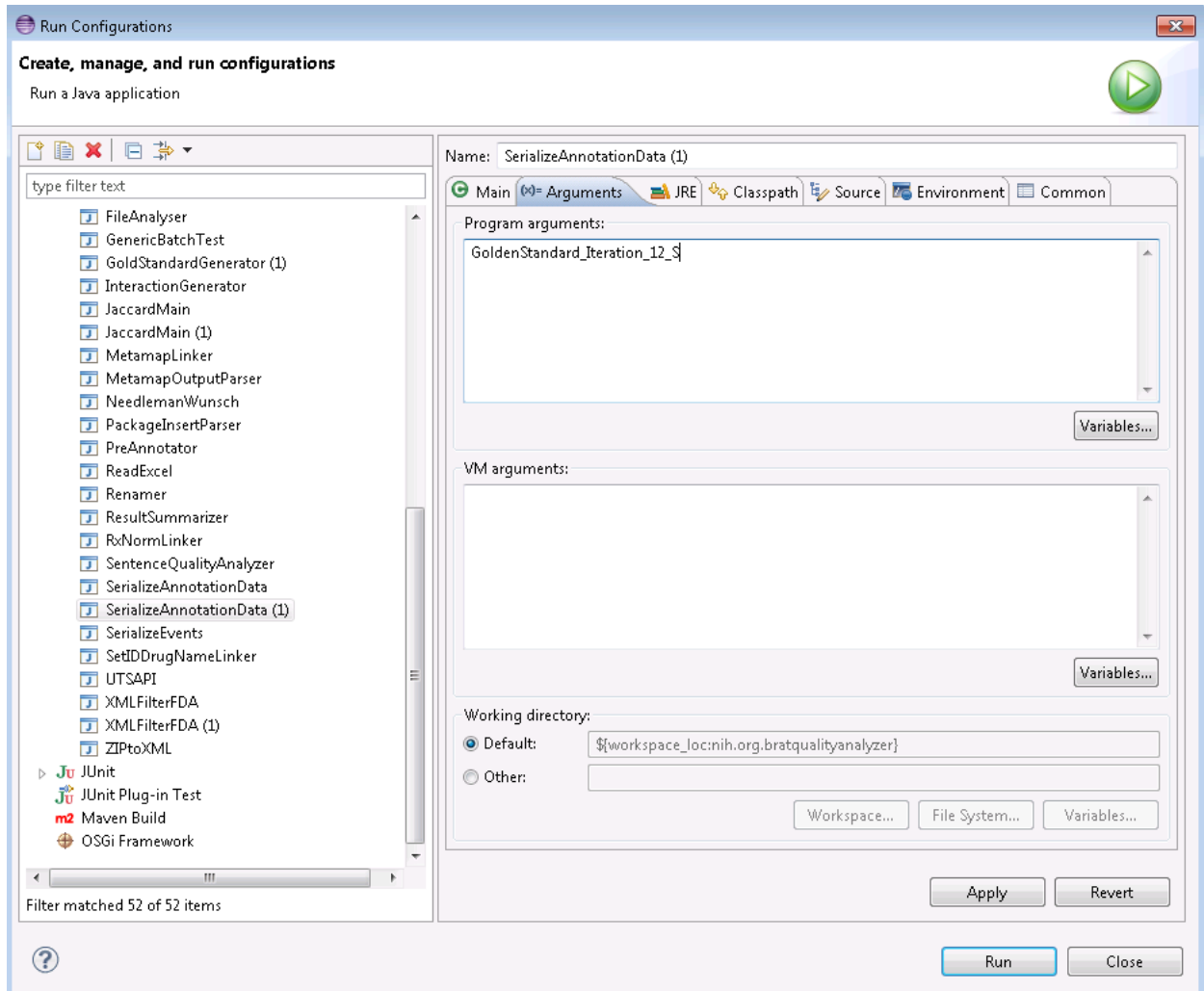


Statistics

- Generates CSV files from the annotations specifically adapted for statistical analysis in R
- `SerializeAnnotationData.java`
 - Serializes every event in a line in the form:
 - Usage: `java [directory of annotated files of the first annotator]`
 - Output:
 - two files in directory called `[directory of annotated files of the first annotator]_csv`
 - `[directory of annotated files of the first annotator]_countR.csv`
 - Format: each line contains the counts of each entity for a file
 - `[directory of annotated files of the first annotator]_eventR.csv`
 - Format: each line an event
- `"ID, " + "File, " + "Caution, " + "Specific, " + "Increase, " + "Decrease, " + "Object, " + "Precipitant, " + "Severity, " + "Context, " + "Evidence"`

- SerializeToCVS.java
 - Usage: `java [directory of annotated files of the first annotator]`
`[directory of annotated files of the second annotator]`

Screenshot for run configuration in Eclipse using test collections provided with the source code (GoldenStandard_Interaction_12*):



- Package: `nih.org.bratqualityanalyzer.sentencesimilarity`
- SentenceQualityAnalyzer.java
 - Usage: `java [directory of annotated files to analyze]`
 - Output (optimized for Excel): list of similar sentence pairs that have different annotation types.
- Output is written into a file called `[directory name of annotated files to analyze]_similar.txt` in a folder called `[directory name of annotated files to analyze]_normalization`:

- File1 – Line number1 – Sentence1 File2 – Line number2 – Sentence2 Differences (list of types that are in only one of them)

- Above examples show 100% similar sentences:

Two patients undergoing desensitizing treatment with hymenoptera venom while receiving ACE inhibitors sustained life-threatening anaphylactoid reactions.

and

Two patients undergoing desensitizing treatment with hymenoptera venom while receiving ACE inhibitors sustained life-threatening anaphylactoid reactions.

- Having a difference in type “Drug” –most probably the span “hymenoptera venom” has been annotated with both Substance and Drug.

In the second example the sentences are very similar and the interaction type is different. Therefore, a manual review is recommended.

Examples

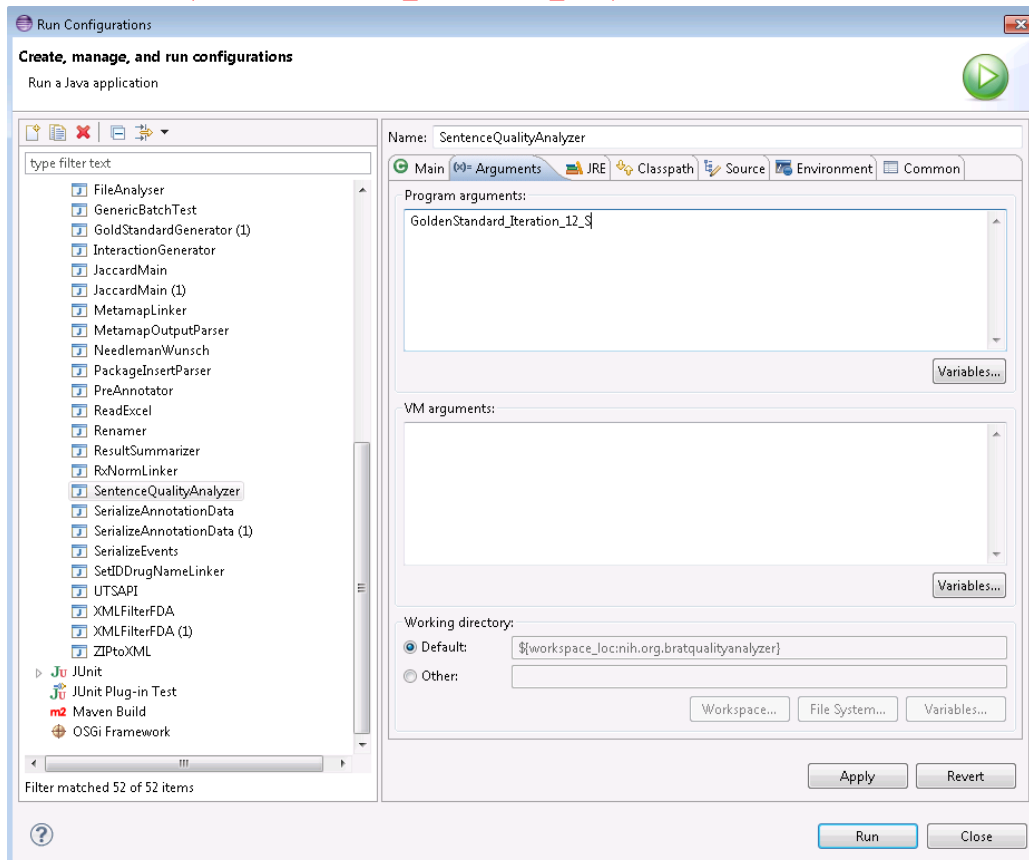
0fc34cd8-86e6-4034-73bd-4263a68ba046	56	Two patients undergoing desensitizing treatment with hymenoptera venom while receiving ACE inhibitors sustained life-threatening anaphylactoid reactions.	94ae6054-b7ae-4212-a567-4f803af8f2c7	54	Two patients undergoing desensitizing treatment with hymenoptera venom while receiving ACE inhibitors sustained life-threatening anaphylactoid reactions.	Drug;
0fc34cd8-86e6-4034-73bd-4263a68ba046	70	Dosage reduction of ALTACE and/or discontinuation of the diuretic may be required.	d44637a7-b980-4bb1-832c-f5a0f73f8259	91	Antidiabetic drugs (e.g., oral agents, insulin): Dosage adjustment of the antidiabetic drug may be required.	Specific Interaction;

- Jaccard similarity is used (set the required threshold in the code), default value is 0.5. Only standard words are considered for the comparison.

- Metamap default semantic types for the identification of biomedical entities to remove from the similarity measure are (this can be changed in file MetamapWebAPI.java) -phsu,dsyn,clnd,sosy,antb,chem,ortf,bdsu,orgf-:
 - `myGenericObj.setField("Batch_Command", "metamap --XMLf -Q 4 -E -J phsu,dsyn,clnd,sosy,antb,chem,ortf,bdsu,orgf");`

Example of transformed sentence: “In combination with d, d administration d to the d of corneal lesions in rats within 5 days.”

Screenshot for run configuration in Eclipse using test collections provided with the source code (GoldenStandard_Iteration_12*):



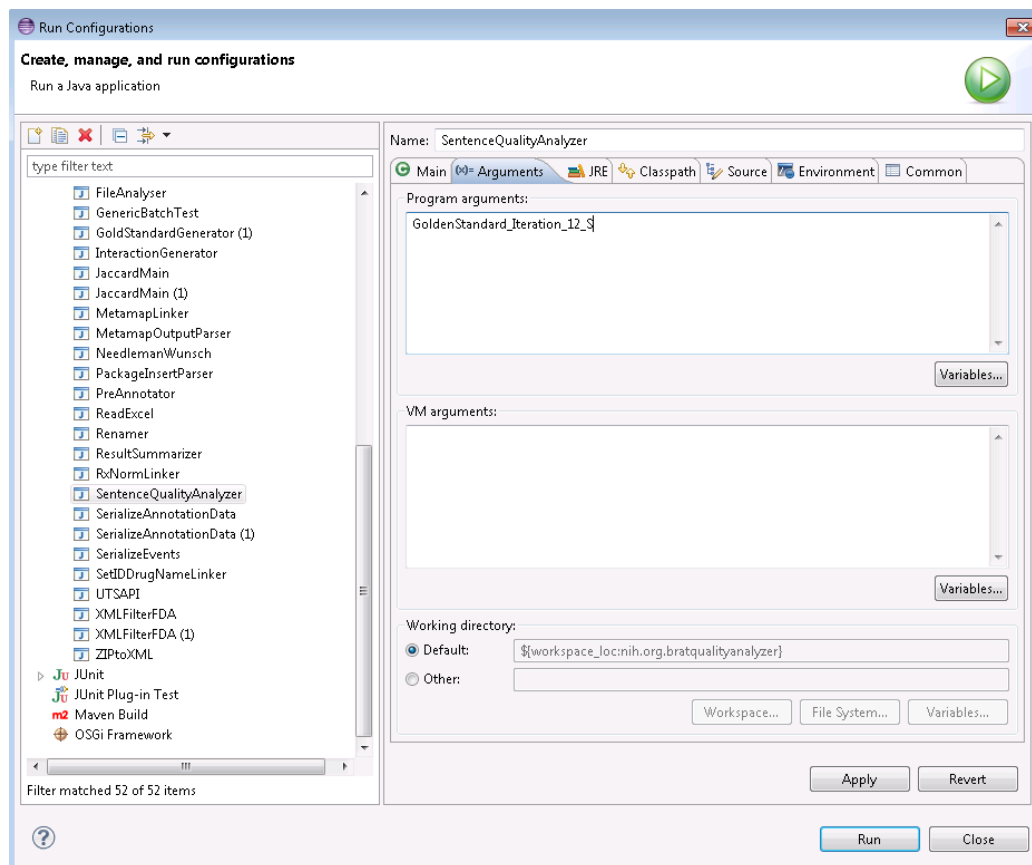
- Package: `nih.org.bratqualityanalyzer.statistics`
- `EntityQualityChecker.java`
 - Usage: `java [directory name of annotated files to analyze]`
 - Output (optimized for Excel): list of files where a given span appears with different type annotations.

The above list shows that spans “potentiation”, “increase the responsiveness” have been annotated with different types. A manual review of these annotations is therefore recommended. Output is written into a file called `[directory name of annotated files to analyze]_entityQuality.txt` in a folder called `[directory name of annotated files to analyze]_normalization`.

POTENTIATION	Specific_Interaction	5562b3f5-8757-11de-8a39-0800200c9a66.ann;	Increase_Interaction	c1302f06-f0f0-407e-817a-52ceb53abb3c.ann; 146d5da7-8c9e-43a2-9a2a-19c0409a6e40.ann; 6e453a4f-34b9-4f07-92b6-bfc28609ce22.ann; eadfe464-720b-4dcd-a0d8-45dba706bd33.ann;
INCREASE THE RESPONSIVENESS	Specific_Interaction	53b6f198-e4a7-4505-9672-efb8c64073ae.ann; 333ee96f-66b6-415f-9ea3-0710b3eee1dd.ann;	Increase_Interaction	0ebcca1e-ca16-4483-429a-aec5dcacd1dc.ann; 146d5da7-8c9e-43a2-9a2a-19c0409a6e40.ann; e2eb7dad-3ea3-439c-dcbb-d1d61aa49dfc.ann;
REDUCTION OF DOSE	Caution_Interaction	77fd7414-85aa-4161-87ff-dbff9e97ee02.ann;	Increase_Interaction	1042fa13-e6af-46b9-8008-6c941f0978b1.ann;
INCREASES IN SERUM LEVELS	Specific_Interaction	2b70cf0c-45be-428f-b396-5001ed4e30fc.ann;	Increase_Interaction	11d14362-8f69-4c30-b487-5d05f6462bd7.ann;
REDUCE DOSE	Caution_Interaction	7c822705-827f-43c7-8c27-35bc31cf6484.ann;	Increase_Interaction	e9481622-7cc6-418a-acb6-c5450daae9b0.ann;

ADDITIVE EFFECT	Specific_Interaction	7885b2a8-be4e-48ab-8113-4e6ab791eb98.ann; 146d5da7-8c9e-43a2-9a2a-19c0409a6e40.ann; 0fed2822-3a03-4b64-9857-c682fcd462bc.ann; 6e453a4f-34b9-4f07-92b6-bfc28609ce22.ann; 4a5762c6-d7a2-4e4c-10b7-8832b36fa5f4.ann; 5562b3f5-8757-11de-8a39-0800200c9a66.ann;	Increase_Interaction	c1302f06-f0f0-407e-817a-52ceb53abb3c.ann; f6f3c339-2c9d-4d07-14a1-6d6c7daf26c6.ann; bb05420c-fd24-4672-9f62-fdd313819287.ann; c2e709d2-96a2-41ea-b5e9-9ce4ccb91f59.ann; fcfc9676-4edc-460e-6199-8161b824908f.ann;
-----------------	----------------------	--	----------------------	---

Screenshot for run configuration in Eclipse using test collections provided with the source code (GoldenStandard_Iteration_12*):



Any questions regarding this software should be addressed to:

Johann Stan, PhD - johann.stan@nih.gov